Usability of various dwell times for eye-gaze-based object selection with eye tracking

Paulus, Yesaya Tommy
Department of Information Systems, Dipa Makassar University

Remijn, Gerard Bastiaan
Faculty of Design, Department of Human Science, Kyushu University

https://hdl.handle.net/2324/7153269

出版情報：Displays. 67, pp.101997-, 2021-04. Elsevier
バージョン：

Displays 67 (2021) 101997

Contents lists available at ScienceDirect

Displays

journal homepage: www.elsevier.com/locate/displa

Usability of various dwell times for eye-gaze-based object selection with eye tracking☆

Yesaya Tommy Paulus [a,*], Gerard Bastiaan Remijn [b]

[a] Department of Information Systems, Dipa Makassar University, Indonesia
[b] Faculty of Design, Department of Human Science, Kyushu University, Fukuoka, Japan

ARTICLE INFO

Keywords:
Eye tracking
Dwell time
Eye-gaze-based input
Object selection

ABSTRACT

This study investigates the usability of various “dwell times” for selecting visual objects with eye-gaze-based input by means of eye tracking. Two experiments are described in which participants used eye-gaze-based input to select visual objects consisting of alphanumeric characters, dots, or visual icons. First, a preliminary experiment was designed to identify the range of dwell time durations suitable for eye-gaze-based object selection. Twelve participants were asked to evaluate, on a 7-point rating scale, how easily they could perform an object-selection task with a dwell time of 250, 500, 1000, or 2000 ms per object. The evaluations showed that a dwell time of 250 ms to around 1000 ms was rated as potentially useful for object selection with eye-gaze-based input. In the following main experiment, therefore, 30 participants used eye tracking to select object sequences from a display with a dwell time of 200, 400, 800, 1000 or 1200 ms per object. Object selection time, object selection success rate, the number of object selection corrections, and dwell time evaluations were obtained. The results showed that the total time necessary to select visual objects (object selection time) increased when dwell time increased, but longer dwell times resulted in a higher object-selection success rate and fewer object selection corrections. Furthermore, regardless of object type, eye-gaze-based object selection with dwell times of 200–800 ms was significantly slower for participants with glasses than for those without glasses. Most importantly, participant evaluations showed that a dwell time of 600 ms per object was easiest to use for eye-gaze-based selection of all three types of visual objects.

1. Introduction

Eye-tracking devices enable the user to select an object on a display, such as a letter, a menu item or a password object, by focusing his/her eye gaze on the object. Eye-gaze-based object selection is regarded as an easy and natural means of human-machine interaction [1], which only needs slight practice [2]. Furthermore, assuming the target object on the display is not too small, eye-gaze-based object selection is also considered as faster than other display-pointing devices, such as a mouse or a stylus [3,4]. Object selection with eye tracking typically requires a certain amount of “dwell time”: to select an object, or to trigger an action by gazing at a button, the user needs to dwell his/her gaze on the object or button for a certain amount of time. Ware and Mikaelian [3] demonstrated that objects can be easily selected by using dwell time only, without the use of extra buttons or other actions. Using dwell time, however, requires some practice by the user. On a display with multiple objects, the user must first reliably identify the target object before an action can be performed on it [5]. Accordingly, when eye gaze is used to identify a target object, the user may unintentionally and inattentively dwell his/her gaze on the wrong object. As a result, this object may even become selected as the target object – a problem that is known as the Midas-Touch problem [6,7,8]. In order to counteract the Midas-Touch problem, developers of eye-gaze-based object selection interfaces typically use a fixed duration of dwell time.

Past research shows that a rather wide range of dwell time durations has been employed. In various eye-typing systems, which employ eye tracking to select characters on an on-screen keyboard, dwell time ranged in between 100 and 1000 milliseconds (ms). For example, a dwell time of 800 ms was used in the “Multi-Tap” and “Symbol-Creator” systems [9]. In the “GazeTalk” system, dwell time was set at 500 ms to type Japanese sentences [10], while dwell times of 400 and 900 ms were used for highlighting and selecting a letter, respectively, to create

☆ This paper was recommended for publication by Prof G Guangtao Zhai.
* Corresponding author.
E-mail addresses: tasyanoah@gmail.com, tasyanoah@dipanegara.ac.id (Y.T. Paulus).

https://doi.org/10.1016/j.displa.2021.101997
Received 6 November 2020; Received in revised form 18 January 2021; Accepted 18 January 2021
Available online 24 February 2021

English words [11]. Similar studies have utilized a dwell time of 600 or 750 ms for sentence typing [12,13]. Finally, in a study in which dwell time could be automatically adjusted to fit the user's needs, dwell time initially ranged from 300 to 1100 ms [14]. Only a few studies have investigated the use of dwell time for other purposes than eye typing. In one study, users were asked to maintain their eye gaze for 1000 ms to select individual password objects on a screen-lock interface [15]. In other studies, object dwell times of 450, 500, or 800 ms, respectively, have been employed for selecting a sequence of characters and faces as a password [16,17,18]. So far, different dwell times thus have been used for eye-gaze-based selection of different types of objects. Furthermore, in all the above studies, dwell time was arbitrarily fixed at one value, without regard of user preferences.

Our aim in the present study is to obtain comparative data of user preferences and capabilities in performing eye-gaze-based object selection with different dwell times. In spite of the wide-spread use of eye-gaze-based object selection for various purposes, to our knowledge, such a study has not been performed yet. First, in a preliminary experiment, participants used a rating scale to indicate their impression of eye-gaze-based object selection under dwell times of 250, 500, 1000, and 2000 ms. This wide range was used to limit the dwell times in the following experiment, in which participants used eye tracking to enter a sequence of 4 objects under 6 different dwell times. The total time necessary to select the correct sequence of objects, object-selection success rate, the number of object selection corrections, and dwell time evaluations were recorded. In both experiments, three types of visual objects were used. The first type consisted of alphanumeric characters, similar to those used in the eye-typing task in most previous research with a dwell time (e.g., Majaranta et al. [11], Bee & André [12], Kurauchi et al. [13]). The second type of visual objects used in this study were patterns of dots, and the third type were visual icons. These types of visual objects are commonly used in recognition-based password systems [19], for example to manually unlock smartphones.

2. Preliminary experiment

The purpose of the experiment was to identify the range of dwell time durations suitable for object selection with eye-gaze-based input. Twelve participants were asked to evaluate, on a 7-point rating scale, how easily they could perform an object-selection task with a dwell time of 250, 500, 1000, and 2000 ms per object. The object-selection task was performed on four grid formations, consisting of 3 × 4, 4 × 3, 4 × 5, and 5 × 4 cells (columns-by-rows, see Fig. 1 below). The dwell time evaluations were obtained for each object type.

2.1. Method

2.1.1. Participants

The participants were 12 students of Kyushu University, Japan, who had normal or corrected-to-normal vision. The age of the participants was in between 21 and 45 years ($M = 28.08$ years, $SD = \pm 7.38$ years). The average height of the participants was 166.25 cm ($SD = \pm 6.68$ cm). The participants were asked to provide written informed consent as to their participation, after they had received an explanation and instructions about the experiment. The participants were paid for their participation. The research was conducted with prior approval of the Ethics Committee of Kyushu University.

2.1.2. Apparatus

Two monitors (20-in, refresh rate 60 Hz) with a resolution of 1600 × 1200 pixels were utilized in this experiment. The first monitor was a Hewlett-Packard LP2065, which was used to present the experiment interface. A Tobii Eye Tracker 4C© device was placed on the lower edge of the monitor, at a height of 133 cm from the ground. The monitor and the eye-tracking device were tilted upwards to two viewing angles of 105° and 120°. These angles were suitable for participants in between 151 and 190 cm in height to register their eye gaze on a very similar eye-tracking system [20]. The second monitor (Lenovo ThinkVision) was used by the experimenter as an interface to control the order of dwell time, grid, and the sequence of 4 or 6 objects that needed to be selected. Both monitors were mounted on a monitor stand, opposite from each other. All experiment interfaces were programmed in Visual Studio C# (2015). Experimental results were saved in a MySQL database. The experiment was performed under room lighting at an illuminance of 124.14 ± 8.23 lux, measured using a TOPCON Illuminance Spectro Meter IM-1000 at the participant's viewing position. Visual object luminance (see below) was measured using a TOPCON Luminance Meter BM-9.

2.1.3. Stimuli

A sequence of 4 or 6 visual objects needed to be selected with eye-gaze-based input, from four different grid formations. These grids consisted of 3 × 4, 4 × 3, 4 × 5, and 5 × 4 cells (columns-by-rows, Fig. 1). The number of objects on the display was the same as the grid density, i.e., 12 objects in 3 × 4 and 4 × 3 grids, and 20 objects in 4 × 5 and 5 × 4 grids.

Three types of objects needed to be selected, in separate trials. The first type were alphanumeric characters consisting of numbers and letters (Fig. 2, left). They were presented in alphabetical order from the top-left to the bottom-right grid cell. For each cell, the numbers and letters were black with a luminance of 0.17 ± 0.01 cd/m^2 on a white background (3.10 ± 0.17 cd/m^2). The second type of objects were dots, which could be sequentially selected to create a shape or a pattern (Fig. 2, middle). Dot patterns are often used as screen locks on smartphones. A single dot on the display was composed of a white circle (3.12 ± 0.21 cd/m^2) with a radius of 47 pixels, which was placed in the middle of a black circle (0.18 ± 0.01 cd/m^2) with a radius of 128 pixels.

The third type of objects were visual icons (Fig. 2, right). The visual icons were placed in a fixed order on one of the four grids, in gray-scale with a luminance range of 0.07–1.18 cd/m^2, against a white background (3.07 ± 0.13 cd/m^2). Typical examples of visual icons are depictions of fruits or animals, of which the essential features necessary for recognition (i.e., the iconic representations) were depicted [21]. For all three types, an object (i.e., alphanumeric character, dot, or icon) was centered in the middle of a grid cell with a size of 128 × 128 pixels, i.e., 4.16 deg × 4.47 deg in visual angle, and every pixel within an object was 0.028

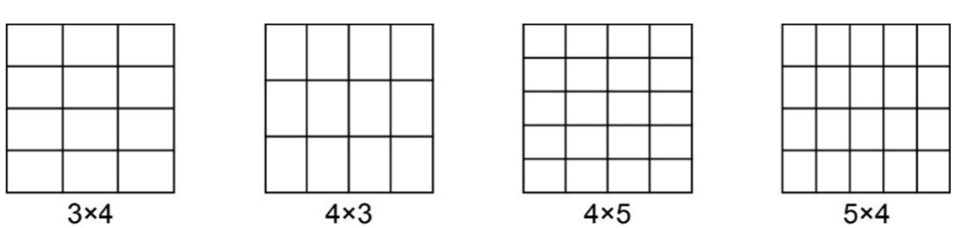

Fig. 1. Schematic impression of the four different grid formations [3 × 4, 4 × 3, 4 × 5, and 5 × 4 cells (columns × rows)] used in the preliminary experiment. Note that regardless of the number of grid cells, the visual objects that needed to be selected with eye-gaze-based input (i.e., alphanumeric characters, dots, or icons) had the same size.

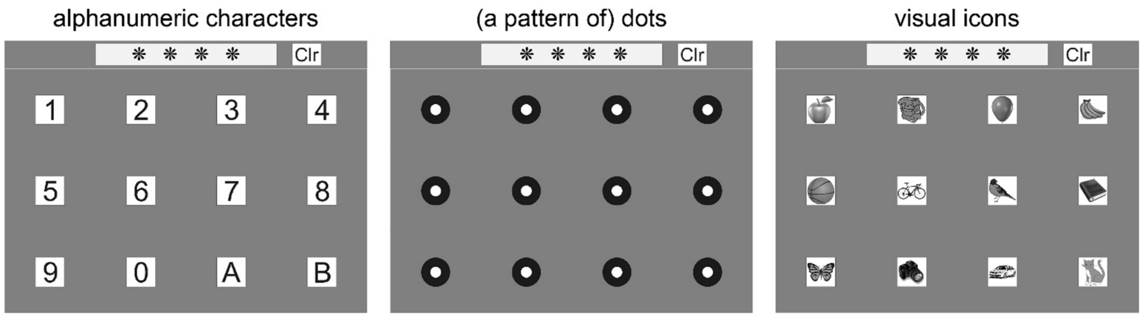

Fig. 2. Examples of the three types of objects that needed to be selected with eye-gaze-based input in this study. A sequence of 4 or 6 visual objects needed to be selected from four grid formations in the preliminary experiment and from a grid formation of 4 × 3 cells in the following experiment.

deg × 0.030 deg in visual angle. The visual object size used in this study was quite adequate for gaze tracking systems as indicated by Ware and Mikaelian [3], who showed that selection performance profoundly dropped for object targets smaller than 1.5 deg in visual angle.

To select a sequence of visual objects (see Procedure below), a screen interface was made with a size of 1600 × 1200 pixels, set against a gray background (1.24 ± 0.05 cd/m^2). In the middle of the upper part of the screen (1600 × 125 pixels) was a text box (800 × 100 pixels), and at the top right was a "Clr" key (128 × 100 pixels). The "Clr" key could be used by the participant to clear incorrectly selected objects one by one. The main part of the screen (1600 × 1075 pixels) displayed the grids and object types. When the participant selected an object on the grid on the main part of the screen, an asterisk would appear in the text box at the upper part of the screen, and a chime sound would be played (pitch corresponding to 670 Hz; 1538 ms) to indicate that an object was selected. All objects on the upper or main part of the screen could be selected by eye gaze under each of the four dwell time durations.

2.1.4. Procedure

When selecting objects with eye-gaze-based input, the participant was standing in the middle in front of the screen at a viewing distance of approximately 49 cm, as indicated by a floor mark. Practically, this viewing distance was close to the minimum operating distance of the eye-tracking device [22], as confirmed in a previous study with a similar eye tracker [23]. The reason the participant performed the task while standing was to simulate a situation in which he/she would use eye tracking to enter an object sequence, such as a password, on an automated teller machine (ATM).

The object-selection task. First, the participant was shown a 4-object or a 6-object sequence, randomly generated for each of the three object types, on a grid that was randomly selected from the four different grids. Thus, the objects that were presented on the screen consisted of a sequence of alphanumeric characters, dots, or visual icons. The participant was then asked to memorize a sequence of 4 objects within a minute and a sequence of 6 objects within two minutes. During this time, the participants could perform memory-maintenance rehearsal of the objects. The appropriate grid was displayed on the screen to assist the participant in memorizing the objects' positions within the grid. After memorizing, the participant was instructed to enter the memorized sequence onto the screen interface by selecting the appropriate objects either with a dwell time of 250, 500, 1000, or 2000 ms. The participant was instructed to use a "Clr" key if he/she had selected an incorrect object. In case the object selection was incorrect, for example, due to the Midas-Touch problem or to selecting objects in the wrong order, the sequence could be attempted up to five times. If the participant failed to enter the correct objects on the fifth attempt, he/she was instructed to try again using a different sequence for the same dwell time, object type and grid.

The evaluation task. After entering the correct objects, the participant was asked to evaluate whether he/she considered the dwell time as easy to use for object selection with eye-gaze-based input. The evaluation was made on a scale between 1 (not easy) and 7 (very easy). The participant used a mouse to make the rating-scale judgments on the screen. The meaning of "easy to use" was defined as the participant's subjective impression of input speed, i.e., estimated time needed to enter the object sequence, and input success, i.e., whether input correction(s) were necessary to enter the objects with eye-gaze-based input.

The experiment was performed with counterbalance in the order of the four dwell time durations. That is, three participants first selected objects with the dwell time of 250 ms, then with the dwell time of 500, 1000 ms, and finally with the dwell time of 2000 ms, for each of the four grids. Next, three participants started with the dwell time of 500 ms, followed by the dwell time of 1000 ms, 2000 ms, and ended with the dwell time of 250 ms. Another three participants started with 1000 ms, followed by 2000 ms, 250 ms, and ended with the dwell time of 500 ms. The remaining three participants started with the dwell time of 2000 ms and ended with the dwell time of 1000 ms. In the same way, the order of object type was also counterbalanced within every four participants. The number of objects in a sequence (4 or 6) was varied as well, for each object type and dwell time.

Before the start of the experiment, the participant needed to register his/her eyes and perform calibration with Tobii Eye Tracker 4C© software. In order to get used to the task, a practice program was prepared in which the participant practiced object selection for a given object type and sequence, on a grid randomly chosen from the four grids. The evaluation task was practiced as well. The preliminary experiment took about 1 h and 30 min, including 40 min for experiment preparation (i.e., instructions, eye registration, and practice).

2.2. Results

The participants evaluated the usability of each dwell time on a rating scale between 1 (not easy) and 7 (very easy). Since the data were not normally distributed, as confirmed by Shapiro-Wilk tests ($p < 0.05$, for four dwell time durations in three object types), non-parametric Friedman tests were performed over the data. If significant, pairwise comparisons using Wilcoxon signed-rank tests with Holm-Bonferroni-correction on the alpha-level based on ranks [$0.05/(m-k + 1)$ were performed, where $m = 6$ indicates the number of pairs and $k = 1, 2, \ldots, m$, indicates the level of ranks], to see which pair(s) of dwell time evaluations were significantly different. Fig. 3 shows the dwell-time evaluations for each object type.

Friedman tests ($df = 3$, $n = 12$) showed that for alphanumeric characters ($\chi^2 = 25.23, p < 0.001$), for dots ($\chi^2 = 18.47, p < 0.001$), and for visual icons ($\chi^2 = 26.84$, $p < 0.001$), the participant evaluations significantly differed. The statistical details of the paired comparisons between dwell time conditions are shown in Table 1. In summary, the results showed that a very long dwell time of 2000 ms was generally judged as not usefull for each of the three object types. For each object type, the dwell time of 500 ms or 1000 ms was judged as significantly more useful. Except for selection of visual icons, the 250-ms dwell time was also evaluated as significantly more useful than the 2000-ms dwell

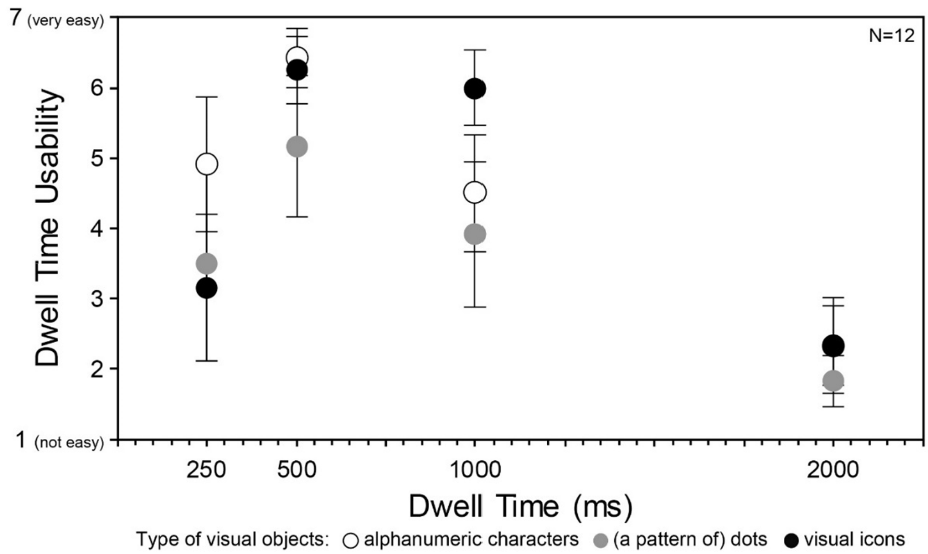

Fig. 3. Average evaluations of dwell-time usability obtained in the preliminary experiment. Evaluations show the user's subjective impression of object input speed and input success, for eye-gaze-based selection of three object types with four different dwell times. Error bars indicate ±95% confidence intervals around the means.

Table 1
Statistical results of the preliminary experiment. Pairwise comparisons of object dwell-time usability for the three object types.

Alphanumeric characters		Dots		Visual icons	
Dwell time (ms) comparison	Z-score	Dwell time (ms) comparison	Z-score	Dwell time (ms) comparison	Z-score
250–1000	−0.63	250–500	**−2.22**†	250–2000	−1.27
250 > 2000	**−2.66****	250–1000	−0.67	500 > 250	**−2.82****
500 > 250	**−2.31***	250–2000	**−2.09**†	500–1000	−0.75
500 > 1000	**−3.10****	500 > 1000	**−2.54***	500 > 2000	**−3.09****
500 > 2000	**−3.10****	500 > 2000	**−2.97****	1000 > 250	**−2.95****
1000 > 2000	**−3.09****	1000 > 2000	**−2.82****	1000 > 2000	**−3.09****

Z-score: Wilcoxon signed rank test value;
>: significantly easier to use;
* $p < 0.05$ after Holm-Bonferroni-correction, ** $p < 0.01$ after Holm-Bonferroni-correction;
† bordered on significance after Holm-Bonferroni-correction.

time. For each object type, the dwell time of 500 ms received the highest evaluation scores. However, since significant differences between the evaluations of the 500-ms dwell time compared to the 250-ms and the 1000-ms dwell time varied with object type (see supplementary Table 1), in the following experiment six dwell times were used in between 200 and 1200 ms. A dwell time of 1200 ms was included to check whether the evaluation of dwell time durations especially around 1000 ms would be similar to either those for 1000 ms or to those for 2000 ms.

Other than dwell time evaluation data, the total time necessary to select the correct sequence of 4 or 6 objects and the number of object selection corrections were also recorded. In short, the data showed that the total selection time for 4 or 6 objects did not vary significantly among the four grid formations. Futhermore, as expected, the total selection time for 6-object sequences was longer than for 4-object sequences without object correction. The data also indicate that with a dwell time of 250 ms, the number of object selection corrections was relatively high, especially for dot and icon objects. Data regarding the total selection time and the number of object selection corrections are included as supplementary materials. Since the preliminary experiment was meant to identify the range of dwell time durations for selecting various object types, further statistical analyses were not performed on these parameters.

3. Experiment

The preliminary experiment showed that participants evaluated a dwell time of 250 ms to 1000 ms as useful for object selection with eye-gaze-based input (Table 1), with differences depending on object type. In the following experiment, therefore, object selection was investigated with dwell times of 200, 400, 600, 800, 1000 and 1200 ms per object. We aimed to investigate the usability of these dwell times for selecting a sequence of 4 visual objects. Three different object types (Fig. 2) were also used in this experiment, but the object-selection task was only performed on a grid of 4 × 3 cells, which is a "horizontal" grid with more columns than rows. Horizontal grids typically enable more efficient object selection with eye-gaze-based input than vertical grids (grids with more rows than columns) [24], which has been corroborated by studies on visual search [25,26,27]. Thirty participants were asked to memorize a 4-object sequence, and to use their eye gaze to enter the sequence of objects onto a user interface. Object-selection time, object-selection success rate, and the number of object selection corrections were obtained and analyzed. Besides entering the sequence of 4 objects onto the user interface, the participants were also required to evaluate the usability of the six dwell time durations.

3.1. Method

3.1.1. Participants

The participants were 30 students of Kyushu University, Japan. Eleven participants had also participated in the preliminary experiment, 11 months earlier. The age of the participants was in between 22 and 47 years ($M = 29.20$ years, $SD = \pm 7.45$ years). Fifteen participants wore glasses, 3 participants wore contact lenses, while 12 participants had uncorrected vision. The average height of the participants was 166.90 cm ($SD = \pm 5.98$ cm). The participants participated on a voluntary basis and provided written informed consent as to their participation after the purpose and procedure of the experiment was explained to them. The research was conducted with prior approval of the Ethics Committee of Kyushu University.

3.1.2. Apparatus, Stimuli, and procedure

The apparatus and the stimuli for this experiment were the same as in the preliminary experiment. The procedure was performed in the same way as in the preliminary experiment. In this experiment, the participant was instructed to select 4-object sequences on a grid of 4 × 3 cells (columns-by-rows) with a dwell time of 200, 400, 600, 800, 1000, and 1200 ms per object. This experiment was also performed with counterbalance in the order of the six dwell time durations within every five participants, and in the order of the three object types within every ten participants. The experiment took about 30–45 min.

3.1.3. Data collections

For every participant, object-selection time, object-selection success rate, the number of object selection corrections, and the dwell-time evaluation data obtained with the rating scale were recorded. Object-selection time was defined as the total time necessary to select the correct sequence of 4 visual objects. Object-selection time was obtained by means of the computer program in 540 trials in total (30 participants $\times$ 6 dwell time durations $\times$ 3 object types). Since both the mean and the standard deviation are particularly sensitive to outliers, the Median Absolute Deviation$_n$ method (MAD$_n$) was used to detect and remove outliers in the data. A removal criterion of 2.5 times the MAD$_n$ above the median was used recursively until no additional outliers were identified [28]. By using this method, no trials were removed. Object-selection success rate was determined by the number of attempts neccesary to perform object selection correctly. The number of object selection corrections for each combination of object type and dwell time was calculated from 150 attempts overall (30 participants $\times$ 5 attempts). In order to ensure a natural attitude towards the task, the participant was not informed about these data collections before the experiment. Finally, the the dwell-time evaluation data obtained with the rating scale were analyzed as in the preliminary experiment.

3.2. Results

3.2.1. Object-selection time

For all three types of visual objects, the results showed that object-selection time increased when dwell time duration increased, as shown in Fig. 4. A regression analysis with a quadratic function was performed in order to examine the relation between object-selection time and dwell time duration. The reason why the quadratic function was used here is to create continuous lines and to estimate object-selection time for any duration among the six dwell times. In this case, x is the dwell time duration (from 200 to 1200 ms), and y is the object-selection time. The regression equations for 4-object alphanumeric, dot, and icon sequences, with r^2 values of 0.998, 0.990, 0.985 ($p \leq 0.002$), respectively, show that the average object-selection time significantly increased with each increase in dwell time duration.

Shapiro-Wilk tests showed that the object-selection time data obtained under each of the six dwell time durations were not normally distributed ($p < 0.05$) for 4-object alphanumeric and dot sequences. Therefore, non-parametric Friedman tests were performed for in order to compare object-selection time between object types. The results (Table 2) show that object-selection time – in case of correct object selection – did not significantly differ between object types for dwell time durations of 600, 800, 1000 and 1200 ms. However, it significantly differed between object types for dwell time durations of 200 ms ($df = 2$, $\chi^2 = 7.20$, $p = 0.027$) and 400 ms ($df = 2$, $\chi^2 = 6.87$, $p = 0.032$). To see which pair(s) of object types showed a significant difference, pairwise comparisons using Wilcoxon signed-rank tests with Holm-Bonferroni-correction on the alpha-level based on ranks [0.05/(m-k + 1) were performed, where m = 3 indicates the number of pairs and k = 1, 2, …, m, indicates the level of ranks]. For a dwell time of 200 ms, participants needed less time to select alphanumeric objects than dot objects (Z = -2.58, $p = 0.010$) or visual icon objects (Z = -2.03, $p = 0.043$). Yet, the

Table 2
Results of the experiment: differences in object-selection time between object types for each dwell time duration.

Dwell Time (ms)				
200		400		600, 800, 1000, 1200
Object type comparison	Z-score	Object type comparison	Z-score	Object-selection time for these dwell time durations did not significantly differ between object types.
Alphanumeric character > Dot	**−2.58***	Alphanumeric character > Dot	**−2.44***	
Alphanumeric character - Visual icon	**−2.03**†	Alphanumeric character > Visual icon	**−2.29***	
Dot - Visual icon	−0.28	Dot - Visual icon	−1.37	

Z-score: Wilcoxon signed rank test value;
>: significantly faster;
* p < 0.05 after Holm-Bonferroni-correction;
† bordered on significance after Holm-Bonferroni-correction.

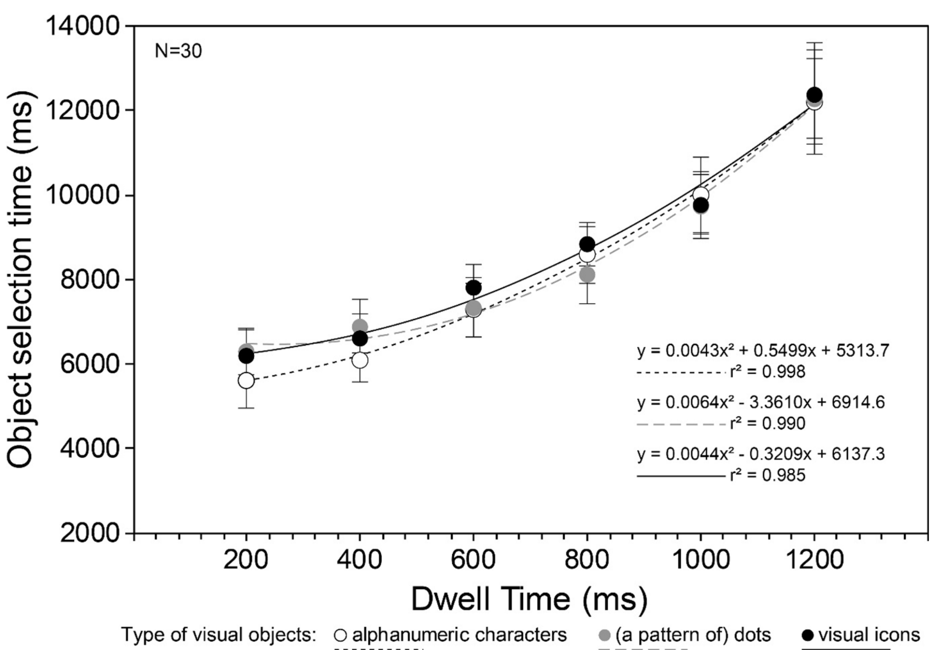

Fig. 4. Results of the experiment: the relation between dwell time per object and the total selection time for 4-object sequences in milliseconds (ms) without object correction. Eye-gaze-based object-selection time was obtained with dwell time durations of 200, 400, 600, 800, 1000 and 1200 ms per object, for three object types. The continuous curves show a quadratic function for each object type through the six dwell time durations. Error bars indicate ±95% confidence intervals around the means.

latter difference did not pass the significance level ($p > 0.025$) after Holm-Bonferroni correction. With a dwell time of 400 ms, participants also significantly needed less time to select alphanumeric objects than dot objects ($Z = -2.44$, $p = 0.015$) or visual icon objects ($Z = -2.29$, $p = 0.022$).

In an additional analysis, we compared object-selection time for each dwell time between participants with glasses against participants without glasses, collapsed over object type. The resulting grand averages (Table 3) showed that object-selection time with dwell times of 200 ms ($Z = -2.92$, $p = 0.004$), 400 ms ($Z = -2.64$, $p = 0.008$), and 800 ms ($Z = -2.25$, $p = 0.024$) was significantly faster for participants without glasses than those with glasses.

3.2.2. Object-selection success rate

Object-selection success was based on the number of attempts necessary to accomplish the task. Data showed that participants generally could correctly select the sequence of four objects from the display on the first attempt, with dwell times of 400–1200 ms. Multiple attempts were usually necessary with a dwell time of 200 ms, but participants often achieved correct object selection at least on the second and third attempt. Therefore, non-parametric binomial tests were performed to see whether the proportion of participants that could complete the selection of four objects correctly in every attempt differed from what was expected. The expected proportion was set with a value of 0.50, as having an "equal chance" of either success or failure.

The results showed that only for a dwell time of 200 ms, the proportion of participants that could select the 4-object sequence correctly at the first attempt, for the three types of objects, was not significantly greater ($p > 0.05$) than the expected proportion. However, the proportion of successful selection at the second attempt (≥ 0.83) and third attempt (≥ 0.93) was significantly greater ($p < 0.001$) than the expected proportion. For dwell times of 400, 600, 800, 1000, and 1200 ms, the proportion of participants that could select the sequence correctly at the first attempt was also significantly greater ($p < 0.001$) than the expected one.

Next to the success rate data, the number of object selection corrections for each dwell time duration and object type was obtained from 150 overall attempts (30 participants $\times$ 5 attempts). Data showed that all participants had accomplished the object-selection tasks before the fifth and final attempt. Indeed with a dwell time of 200 ms, however, the number of object selection corrections was relatively high for the three object types: 11%, 13%, and 14%, respectively, for alphanumeric characters, dots, and visual icons. The number of object selection corrections strongly decreased – often to zero – as dwell time duration increased. Because of the relatively low number of object selection corrections overall, further statistical analyses were not performed.

3.2.3. Dwell-time evaluations

Similar to the preliminary experiment, the participants also evaluated the usability of each dwell time on a 7-point rating scale between 1 (not easy) and 7 (very easy) in this experiment. Participants' evaluations were based on their subjective impression of how fast (estimated time needed) and with how few corrections they thought they had entered the object sequence. A regression analysis with a quadratic function was performed to establish continuous curves between the usability ratings and dwell time duration (Fig. 5). The regression equations for alphanumeric, dot and icon objects, with r^2 values of 0.917, 0.966, 0.986 ($p < 0.025$), respectively, suggest that the continuous curves fit very well to the data. Furthermore, the equations also could estimate the optimum dwell time for each object type. According to the regression equations, the optimum estimated dwell time for alphanumeric objects was 575 ms, for dot objects it was 625 ms, and for visual icons it was 650 ms.

The statistical analyses, performed in a similar way as in the preliminary experiment, are shown in Table 4. Friedman tests ($df = 5$, n = 30) for alphanumeric characters ($\chi^2 = 83.29$, $p < 0.001$), for dot patterns ($\chi^2 = 71.07$, $p < 0.001$), and for visual icons ($\chi^2 = 86.59$, $p < 0.001$) showed significantly different evaluations between dwell time durations. Table 4 shows that the usability of the 600-ms dwell time, around which the optima were estimated, was indeed significantly better than the dwell times of 200 ms, 1000 ms, and 1200 ms for all object types. For object selection of alphanumeric characters, the usability of the 600-ms dwell time was rated significantly higher than that of the 800-ms dwell time, but not of the 400-ms dwell time. For dots, no significant differences were found between the evaluations of the 400-ms, 600-ms, and the 800-ms dwell times. For visual icons, the 600-ms dwell time was close to significantly more useable than the 400-ms dwell time, but not the 800-ms dwell time. In short, the statistics thus confirm that for any object type the 600-ms dwell time was rated as having good usability overall.

Table 3
Results of the experiment: differences in average object-selection time (ms) of each dwell time duration for participants who wore glasses against participants who did not wear glasses.

Dwell Time (ms)	Average object-selection time (ms) ± 95% Confidence Intervals		Z-score
	Participants without glasses (n = 15)	Participants with glasses (n = 15)	
200	5555.63 ± 430.11	6538.15 ± 502.74	**−2.918****
400	6032.48 ± 510.19	7056.54 ± 380.75	**−2.636****
600	7061.25 ± 508.70	7886.85 ± 483.31	**−1.925**†
800	8008.22 ± 521.46	9037.53 ± 453.51	**−2.252***
1000	9519.52 ± 657.60	10173.34 ± 610.70	−1.123
1200	11687.34 ± 933.62	12894.45 ± 856.27	−1.800

Z-score: Wilcoxon signed rank test value;
* p < 0.05, ** p < 0.01, † bordered on significance.

4. Discussion and conclusion

In this study, we aimed to investigate the usability of various dwell times for selecting a sequence of visual objects on a display by using eye tracking. To achieve this aim, two experiments were conducted in which three different visual object types were used. First, a preliminary experiment was performed to estimate the range of dwell time durations that can be used for selecting different types of visual objects. Based on the results, in the following experiment, object selection from a display with eye-gaze-based input was investigated with object dwell times of 200, 400, 600, 800, 1000, or 1200 ms. This experiment aimed to investigate the usability of these dwell times for selecting a sequence of 4 objects on a grid formation of 4 $\times$ 3 cells. Thirty participants were asked to memorize a 4-object sequence, and to use their eye gaze to enter the sequence of objects onto a user interface. The object-selection time, object-selection success rate and the number of object selection corrections were obtained and analyzed. Besides entering the sequence of objects onto the user interface, the participants were also required to evaluate the usability of the six dwell time durations. The results of the experiment showed the following.

Firstly, we found that the number of object selection corrections decreased with every increase in dwell time duration. Most object selection corrections were made with the dwell time of 200 ms for each object type. This dwell time is the minimal border of the human visual fixation range, which is 200–600 ms [4,29]. The fixation range is the time necessary to stabilize the eyes to fixate on something in the visual field. As mentioned in the Introduction, it is likely that the Midas-Touch problem occurred with the 200-ms dwell time [6,7]. With a relatively short dwell time, participants may have unintentionally selected objects while still scanning the display to identify potential target objects to form the correct sequence. These unwanted selections needed to be corrected, and hence the number of selection corrections was relatively high under a 200-ms dwell time per object. In reading tasks, lexical activation and recognition generally require fixation durations on

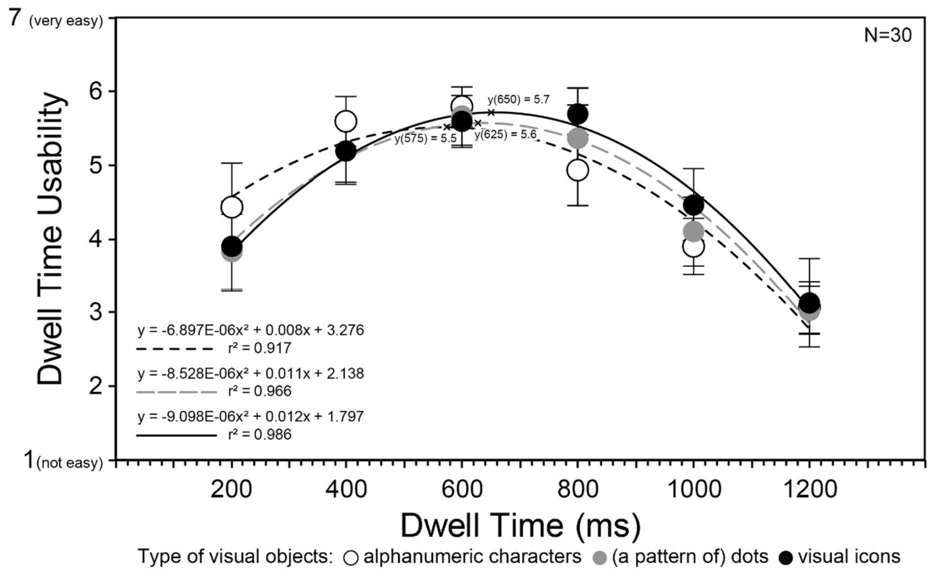

Fig. 5. Results of the experiment: average evaluations of the usability of each dwell time duration (ms) for eye-gaze-based selection of three object types. The continuous curves show a quadratic function for each maximum point and for each object type through the six dwell time durations. The crosses show the estimated maximum points of the dwell time usability ratings for each object type. Error bars indicate ±95% confidence intervals around the means.

Table 4
Results of the experiment: pairwise comparisons of object dwell-time usability for three object types.

Alphanumeric characters		Dots		Visual icons	
Dwell time (ms) comparison	Z-score	Dwell time (ms) comparison	Z-score	Dwell time (ms) comparison	Z-score
200–800	−1.27	200–1000	−1.08	200–1000	−1.53
200–1000	−1.56	200–1200	**−2.50**†	200–1200	**−2.08**†
200 > 1200	**−3.34****	400 > 200	**−3.50****	400 > 200	**−3.59****
400 > 200	**−3.61****	400–600	−1.89	400–600	**−2.03**†
400–600	−1.13	400–800	−0.74	400–800	−1.91
400–800	−1.93	400 > 1000	**−3.05****	400–1000	**−2.07**†
400 > 1000	**−4.27****	400 > 1200	**−4.31****	400 > 1200	**−4.48****
400 > 1200	**−4.65****	600 > 200	**−3.91****	600 > 200	**−3.57****
600 > 200	**−3.44****	600–800	−1.37	600–800	−0.65
600 > 800	**−3.13****	600 > 1000	**−3.80****	600 > 1000	**−3.51****
600 > 1000	**−4.60****	600 > 1200	**−4.75****	600 > 1200	**−4.83****
600 > 1200	**−4.82****	800 > 200	**−3.30****	800 > 200	**−3.84****
800 > 1000	**−3.96****	800 > 1000	**−4.08****	800 > 1000	**−4.08****
800 > 1200	**−4.70****	800 > 1200	**−4.82****	800 > 1200	**−4.85****
1000 > 1200	**−4.13****	1000 > 1200	**−3.97****	1000 > 1200	**−4.27****

Z-score: Wilcoxon signed rank test value;
>: significantly easier to use;
** $p < 0.01$ after Holm-Bonferroni-correction;
† bordered on significance after Holm-Bonferroni-correction.

average in between 200 and 300 ms, whereas in scene perception and visual search tasks, the fixation duration is approximately in between 180 and 330 ms [30]. Thus, with a short dwell time of 200 ms, the visual object recognition process is still in progress.

Secondly, object-selection success rates, i.e., the total attempts necessary to complete object selection tasks correctly with each dwell time, showed that regardless of object type the majority of the participants (≥90%) accomplished eye-gaze-based object selection at the first attempt with dwell times of 400, 600, 800, 1000, and 1200 ms. By contrast, most participants needed more attempts to correctly complete object selection with a 200-ms dwell time. Again, it is most likely that the Midas-Touch problem occurred with the 200-ms dwell time [6,7], and that participants needed to make more corrections. Nevertheless, even with the 200-ms dwell time, a high percentage of trials could be completed at the 2nd attempt (≥83%) or at the 3rd attempt (≥93%). Future research is necessary to clarify this issue further, but it is plausible that performance improved with practice.

Thirdly, the total selection time of 4-object sequences varied depending on the type of objects to be selected and the use of glasses. Object-sequence selection time in case of correct object selection significantly increased with every increase in dwell time duration (Fig. 4). This is in line with a previous study on object-selection time using eye tracking [31]. For any dwell-time duration between 200 and 1200 ms, object-sequence selection time could be predicted quite well with a quadratic function. Here we found that selection of a sequence of four characters was significantly faster than dot and icon object selection with dwell times of 200 and 400 ms (Table 2). Participants most likely were already familiar with selection of alphanumeric characters from a display, but not with dot and visual icon object selection. This might have facilitated object identification and enabled them to select the correct sequence even with short dwell times. Furthermore, regardless of object type, eye-gaze-based object selection with dwell times of 200–800 ms was significantly faster for participants without glasses than for those with glasses (Table 3). A possible explanation for this finding is that for participants with prescription glasses, the registered eye-gaze point often can drift far off from the real eye-gaze point [20,32].

Finally, the important finding in the dwell-time evaluations was that the usability of the 600-ms dwell time was rated high overall for eye-gaze-based selection of each of the three object types (Fig. 5). The regression analysis with quadratic functions also showed that the optimum estimated dwell time for each visual object was around 600 ms. Besides that, the 400-ms dwell time (for alphanumeric characters) and the 800-ms dwell time (for visual icons) were also rated as relatively usable. In line with a related study on retrieving information from a display [33], the longer dwell times of 1000 and 1200 ms were not considered as very usable. Many participants mentioned that they thought the interface or the eye-tracking system were not working well when using these longer dwell times, although very few object selection corrections were necessary with these dwell times. As for the shortest dwell time of 200 ms, the overall ratings were also relatively low, most likely due to the number of incorrect object selections and, hence, corrections. To avoid this Midas-touch problem, and given the participant evaluations, the relatively high object-selection success rate, and the relatively low object-selection time to select a sequence of objects, the present study thus suggests that a generic object dwell time of 600 ms is recommendable for eye-gaze-based input.

In future research, some points will need further investigation. One issue open to investigation is whether older users would have the same dwell time usability evaluations as younger users. It is known that the visual area in which information can be obtained within one eye fixation reduces in size as a function of age [34]. Furthermore, visual search accuracy is generally also affected by age [35,36]. Another issue as suggested here is the effect of practice on object-selection time, the selection success rate, and dwell-time evaluation. For example, it is known that participants may get used to searching visual objects on a screen after a few days of practice [37,38]. As a result, visual search time may decrease and visual search accuracy may increase or at least become consistent. It is thus important to investigate the effects of practice, especially with dot and visual icon objects. There is a possibility that the Midas-Touch problem will also become less with increasing user experience, which would suggest that an object dwell time of even less than 600 ms – e.g., from around 400 to around 500 ms – may become feasible in certain systems.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

We would like to thank the participants for their time and efforts. This study is supported by the Q-DAI Jump Program (FY 2017-2020) and JSPS KAKENHI Grant Number 20K20866.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.displa.2021.101997.

References

[1] P. Majaranta, K.-J. Räihä, Twenty Years of Eye Typing: Systems and Design Issues, in: Proc. Symp. Eye Track. Res. Appl. - ETRA '02, ACM Press, New York, New York, USA, 2002, p. 15. https://doi.org/10.1145/507072.507076.

[2] D.M. Stampe, E.M. Reingold, Selection By Looking: A Novel Computer Interface And Its Application To Psychological Research, in: J. Geotech. Geoenvironmental Eng. ASCE, 1995, pp. 467–478. https://doi.org/10.1016/S0926-907X(05)80039-X.

[3] C. Ware, H.H. Mikaelian, An evaluation of an eye tracker as a device for computer input2, ACM SIGCHI Bull. 17 (1987) 183–188, https://doi.org/10.1145/30851.275627.

[4] L.E. Sibert, R.J.K. Jacob, Evaluation of eye gaze interaction, in: Proc. SIGCHI Conf. Hum. Factors Comput. Syst. - CHI '00, ACM Press, New York, New York, USA, 2000, pp. 281–288. https://doi.org/10.1145/332040.332445.

[5] M.F. Land, S. Furneaux, The knowledge base of the oculomotor system, Philos. Trans. R. Soc. London. Ser. B Biol. Sci. 352 (1997) 1231–1239, https://doi.org/10.1098/rstb.1997.0105.

[6] R.J.K. Jacob, The use of eye movements in human-computer interaction techniques: what you look at is what you get, ACM Trans. Inf. Syst. 9 (1991) 152–169, https://doi.org/10.1145/123078.128728.

[7] B.B. Velichkovsky, M.A. Rumyantsev, M.A. Morozov, New solution to the midas touch problem: identification of visual commands via extraction of focal fixations, Procedia Comput. Sci. 39 (2014) 75–82, https://doi.org/10.1016/j.procs.2014.11.012.

[8] Y.K. Meena, H. Cecotti, K. Wong-Lin, G. Prasad, A multimodal interface to resolve the Midas-Touch problem in gaze controlled wheelchair, in: 2017 39th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc., IEEE, 2017, pp. 905–908. https://doi.org/10.1109/EMBC.2017.8036971.

[9] D. Miniotas, O. Spakov, G. Evreinov, Symbol creator : an alternative eye-based text entry technique with low demand for screen space, Proc. Human-Computer Interact. - INTERACT'03 (2003) 137–143.

[10] J.P. Hansen, A.S. Johansen, D.W. Hansen, K. Itoh, S. Mashino, Command Without a Click: Novice User Studies of Dwell Time Typing by Mouse and Eye-Gaze Selections, Human-Computer Interact. – INTERACT'03. (2003) 121–128.

[11] P. Majaranta, I.S. MacKenzie, A. Aula, K.-J. Räihä, Auditory and visual feedback during eye typing, in: CHI '03 Ext. Abstr. Hum. Factors Comput. Syst. - CHI '03, ACM Press, New York, New York, USA, 2003, p. 766. https://doi.org/10.1145/765978.765979.

[12] N. Bee, E. André, Writing with Your Eye: A Dwell Time Free Writing System Adapted to the Nature of Human Eye Gaze, in: Percept, in: Multimodal Dialogue Syst, Springer, Berlin Heidelberg, Berlin, Heidelberg, 2008, pp. 111–122, https://doi.org/10.1007/978-3-540-69369-7_13.

[13] A. Kurauchi, W. Feng, A. Joshi, C. Morimoto, M. Betke, EyeSwipe: Dwell-free Text Entry Using Gaze Paths, in: Proc. 2016 CHI Conf. Hum. Factors Comput. Syst. - CHI '16, ACM Press, New York, New York, USA, 2016, pp. 1952–1956. https://doi.org/10.1145/2858036.2858335.

[14] O. Spakov, D. Miniotas, On-line adjustment of dwell time for target selection by gaze, in: Proc. Third Nord. Conf. Human-Computer Interact. - Nord. '04, ACM Press, New York, New York, USA, 2004, pp. 203–206. https://doi.org/10.1145/1028014.1028045.

[15] A. Maeder, C. Fookes, S. Sridharan, Gaze based user authentication for personal computer applications, in: Proc. 2004 Int. Symp. Intell. Multimedia, Video Speech Process. 2004., IEEE, 2005, pp. 727–730. https://doi.org/10.1109/ISIMP.2004.1434167.

[16] M. Kumar, T. Garfinkel, D. Boneh, T. Winograd, Reducing Shoulder-surfing by Using Gaze-based Password Entry, in: SOUPS '07 Proc. 3rd Symp. Usable Priv. Secur., 2007, pp. 13–19. https://cups.cs.cmu.edu/soups/2007/proceedings/p13_kumar.pdf.

[17] A. De Luca, R. Weiss, H. Drewes, Evaluation of eye-gaze interaction methods for security enhanced PIN-entry, in: Proc. 19th Australas. Conf. Comput. Interact. Entertain. User Interfaces - OZCHI '07, ACM Press, New York, New York, USA, 2007, pp. 199–202. https://doi.org/10.1145/1324892.1324932.

[18] P. Dunphy, A. Fitch, P. Olivier, Gaze-contingent passwords at the ATM, in: Proc. 4th Conf. Commun. by Gaze Interact. – COGAIN, COGAIN NoE, 2008, pp. 59–62. http://homepages.cs.ncl.ac.uk/p.m.dunphy/COGAIN2008.pdf.

[19] R. Biddle, S. Chiasson, P.C. van Oorschot, Graphical passwords: learning from the first twelve years, ACM Comput. Surv. 44 (2012) 1–41, https://doi.org/10.1145/2333112.2333114.

[20] Y.T. Paulus, G.B. Remijn, Y.K.H. Syn, C. Hiramatsu, The use of glasses during registration into a low-cost eye tracking device under different lighting conditions, in: 2017 2nd Int. Conf. Autom. Cogn. Sci. Opt. Micro Electro--Mechanical Syst. Inf. Technol., IEEE, 2017, pp. 59–64. https://doi.org/10.1109/ICACOMIT.2017.8253387.

[21] C. Gatsou, A. Politis, D. Zevgolis, From icons perception to mobile interaction, 2011 Fed Conf. Comput. Sci. Inf. Syst. FedCSIS. (2011) 705–710.

[22] Tobii Eye Tracking Support, Eye Tracker 4C, Popular topics - HELP.TOBII.COM, 2017. What's the difference between Tobii Eye Tracker 4C and Tobii EyeX?. https://help.tobii.com/hc/en-us/articles/212814329-What-s-the-difference-between-Tobii-Eye-Tracker-4C-and-Tobii-EyeX-/ (accessed 10 August 2019).

[23] Y.T. Paulus, C. Hiramatsu, Y.K.H. Syn, G.B. Remijn, Measurement of viewing distances and angles for eye tracking under different lighting conditions, in: 2017 2nd Int. Conf. Autom. Cogn. Sci. Opt. Micro Electro--Mechanical Syst. Inf. Technol., IEEE, 2017, pp. 54–58. https://doi.org/10.1109/ICACOMIT.2017.8253386.

[24] Y.T. Paulus, G.B. Remijn, What kind of grid formations and password formats are useful for password authentication with eye-gaze-based input? J. Ergon. 09 (2019) https://doi.org/10.35248/2165-7556.19.9.249.

[25] Andrew T. Duchowski, Industrial engineering and human factors, in: Springer, Eye Tracking Methodology: Theory and Practice, second ed., Verlag London, 2007, pp. 242–260.

[26] H. Ojanpää, R. Näsänen, I. Kojo, Eye movements in the visual search of word lists, Vision Res. 42 (2002) 1499–1512.

[27] R.S. Goonetilleke, W.C. Lau, H.M. Shih, Visual search strategies and eye movements when searching Chinese character screens, Int. J. Hum. Comput. Stud. 57 (2002) 447–468, https://doi.org/10.1006/ijhc.2002.1027.

[28] C. Leys, C. Ley, O. Klein, P. Bernard, L. Licata, Detecting outliers: do not use standard deviation around the mean, use absolute deviation around the median, J. Exp. Soc. Psychol. 49 (2013) 764–766, https://doi.org/10.1016/j.jesp.2013.03.013.

[29] V. Cantoni, C. Galdi, M. Nappi, M. Porta, D. Riccio, GANT: Gaze analysis technique for human identification, Pattern Recognit. 48 (2015) 1027–1038, https://doi.org/10.1016/j.patcog.2014.02.017.

[30] K. Rayner, The 35th Sir Frederick Bartlett Lecture: Eye movements and attention in reading, scene perception, and visual search, Q. J. Exp. Psychol. 62 (2009) 1457–1506, https://doi.org/10.1080/17470210902816461.

[31] A.M. Penkar, C. Lutteroth, G. Weber, Designing for the Eye – Design Parameters for Dwell in Gaze Interaction, in: Proc. 24th Aust. Comput. Interact. Conf. - OzCHI '12, ACM Press, New York, New York, USA, 2012, pp. 479–488. https://doi.org/10.1145/2414536.2414609.

[32] Y.T. Paulus, Research on eye-gaze-based input for visual password authentication, Kyushu University, Fukuoka, Japan, 2020 http://hdl.handle.net/2324/4060175 (accessed November 5, 2020).

[33] P. Majaranta, A. Aula, K.-J. Räihä, Effects of feedback on eye typing with a short dwell time, in: Proc. Eye Track. Res. Appl. Symp. Eye Track. Res. Appl. - ETRA'2004, ACM Press, New York, New York, USA, 2004, pp. 139–146. https://doi.org/10.1145/968363.968390.

[34] K.K. Ball, B.L. Beard, D.L. Roenker, R.L. Miller, D.S. Griggs, Age and visual search: expanding the useful field of view, J. Opt. Soc. Am. A. 5 (1988) 2210–2219, https://doi.org/10.1364/JOSAA.5.002210.

[35] D.J. Madden, L.R. Gottlob, P.A. Allen, Adult age differences in visual search accuracy: attentional guidance and target detectability, Psychol. Aging. 14 (1999) 683–694, https://doi.org/10.1037/0882-7974.14.4.683.

[36] S.C. Lee, Y.W. Kim, Y.G. Ji, Effects of visual complexity of in-vehicle information display: age-related differences in visual search task in the driving context, Appl. Ergon. 81 (2019) 102888, https://doi.org/10.1016/j.apergo.2019.102888.
[37] F. Baluch, L. Itti, Training top-down attention improves performance on a triple-conjunction search task, PLoS One 5 (2010), https://doi.org/10.1371/journal.pone.0009127.
[38] K. Clark, L.G. Appelbaum, B. van den Berg, S.R. Mitroff, M.G. Woldorff, Improvement in visual search with practice: mapping learning-related changes in neurocognitive stages of processing, J. Neurosci. 35 (2015) 5351–5359, https://doi.org/10.1523/JNEUROSCI.1152-14.2015.